AD-A092 928 GEORGE WASHINGTON UNIV WASHINGTON DC INST FOR MANAGE--ETC F/G 12/1
IMPROVEMENT OF KERNEL ESTIMATORS OF THE FAILURE RATE FUNCTION U--ETC(U)
FEB 80 N D SINGPURWALLA, M WONG N00014-77-C-0263
UNCLASSIFIED SERIAL-T-415 NL

1 OF 1
AD
A092928

END
DATE
FILMED
1 81
DTIC

5

LEVEL II (12)

AD A092928

IMPROVEMENT OF KERNEL ESTIMATORS OF THE FAILURE RATE FUNCTION USING THE GENERALIZED JACKKNIFE

by

Nozer D. Singpurwalla
Man-Yuen Wong

THE
GEORGE
WASHINGTON
UNIVERSITY

STUDENTS FACULTY STUDY R
ESEARCH DEVELOPMENT FUT
URE CAREER CREATIVITY CO
MMUNITY LEADERSHIP TECH
NOLOGY FRONTIER DESIGN
ENGINEERING APP ENC
GEORGE WASHIN UNIV

SCHOOL OF ENGINEERING
AND APPLIED SCIENCE

LEVEL II (12)

IMPROVEMENT OF KERNEL ESTIMATORS OF
THE FAILURE RATE FUNCTION USING
THE GENERALIZED JACKKNIFE

by

Nozer D. Singpurwalla
Man-Yuen Wong

Serial T-415
14 February 1980

The George Washington University
School of Engineering and Applied Science
Institute for Management Science and Engineering

Contract N00014-77-C-0263
Project NR-042-372
Office of Naval Research

NONE

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER T-415	2. GOVT ACCESSION NO. AD-A092 928	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) IMPROVEMENT OF KERNEL ESTIMATORS OF THE FAILURE RATE FUNCTION USING THE GENERALIZED JACKKNIFE.		5. TYPE OF REPORT & PERIOD COVERED SCIENTIFIC
		6. PERFORMING ORG. REPORT NUMBER T-415
7. AUTHOR(s) NOZER D. SINGPURWALLA, MAN-YUEN WONG		8. CONTRACT OR GRANT NUMBER(s) N00014-77-C-0263
9. PERFORMING ORGANIZATION NAME AND ADDRESS THE GEORGE WASHINGTON UNIVERSITY INSTITUTE FOR MANAGEMENT SCIENCE & ENGINEERING WASHINGTON, DC 20052		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS OFFICE OF NAVAL RESEARCH CODE 434 ARLINGTON, VA 22217		12. REPORT DATE 14 February 1980
		13. NUMBER OF PAGES 41
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)		15. SECURITY CLASS. (of this report) NONE
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

APPROVED FOR PUBLIC SALE AND RELEASE: DISTRIBUTION IS UNLIMITED.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

FAILURE RATE FUNCTION
KERNEL ESTIMATION
RATES OF CONVERGENCE OF MEAN SQUARE ERROR
JACKKNIFING

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

In this paper we explore methods by which the rate of convergence of the bias and the mean square error of kernel estimators of the failure rate function can be improved. We show that if the kernel is not restricted to be nonnegative, and is suitably chosen, then the bias contribution to the asymptotic mean square error can be eliminated to any required order, and the rate of convergence of the asymptotic mean square error can be brought

(continued)

DD FORM 1473 1 JAN 73 EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-014-6601

NONE

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

406743

NONE

SECURITY CLASSIFICATION OF THIS PAGE(When Data Entered)

20. Abstract (continued)

as close to n^{-1} as is desired. The generalized jackknife method of combining estimators is shwon to be an adequate procedure which leads us to this goal.

NONE

SECURITY CLASSIFICATION OF THIS PAGE(When Data Entered)

THE GEORGE WASHINGTON UNIVERSITY
School of Engineering and Applied Science
Institute for Management Science and Engineering

Abstract
of
Serial T-415
14 February 1980

IMPROVEMENT OF KERNEL ESTIMATORS OF THE FAILURE RATE FUNCTION USING THE GENERALIZED JACKKNIFE

by

Nozer D. Singpurwalla
Man-Yuen Wong

In this paper we explore methods by which the rate of convergence of the bias and the mean square error of kernel estimators of the failure rate function can be improved. We show that if the kernel is not restricted to be nonnegative, and is suitably chosen, then the bias contribution to the asymptotic mean square error can be eliminated to any required order, and the rate of convergence of the asymptotic mean square error can be brought as close to n^{-1} as is desired. The generalized jackknife method of combining estimators is shown to be an adequate procedure which leads us to this goal.

Research Supported by
Contract N00014-77-C-0263
Project NR-042-372
Office of Naval Research

THE GEORGE WASHINGTON UNIVERSITY
School of Engineering and Applied Science
Institute for Management Science and Engineering

IMPROVEMENT OF KERNEL ESTIMATORS OF THE FAILURE RATE FUNCTION USING THE GENERALIZED JACKKNIFE

by

Nozer D. Singpurwalla
Man-Yuen Wong

1. Introduction and Summary

The failure rate function is one of the most important parameters in reliability theory. Of the several methods for estimating the failure rate that have been proposed, those based upon weighting functions or "kernels" are quite common. These kernels have, among other things, the following two features:

(i) they are nonnegative, and

(ii) they are absolutely integrable in $(-\infty,\infty)$;

such kernels are called L^1 kernels. Watson and Leadbetter (1964a, 1964b) show that the kernel estimators of the failure rate at some point x_0 are usually biased; the bias converges to zero as the sample size increases to infinity.

Our motivation for undertaking the research reported here is to explore ways in which we can reduce the bias, and improve upon the rate of convergence of the mean square error (MSE). Our conclusion is that if the failure rate function is sufficiently "smooth," and if the kernel is suitably chosen, then the bias contribution to the asymptotic MSE can in principle be eliminated to any desired order, and that the

rate of convergence of the asymptotic MSE can be brought as close to n^{-1} as is desired. However, in order to achieve these properties, the nonnegativity and/or the absolute integrability restrictions on the kernels will have to be relaxed.

An interesting (though not surprising) aspect of our work is that the desired kernels can be constructed by the "generalized jackknife" (GJ) method of Gray and Schucany [cf. Schucany and Sommers (1977)]. Viewed alternatively, this means that if we use the GJ on two kernel estimators of the failure rate, with each estimator based upon a nonnegative L^1 kernel then, this is equivalent to directly producing a kernel estimator using a kernel not restricted to be nonnegative. Finally, we conjecture that using the GJ indefinitely is equivalent to producing a kernel estimator using a kernel which does not satisfy both (i) and (ii) above. Since kernel estimators are also used in density estimation, and the estimation of the power spectrum, the above results should be of a wider interest.

The organization of our paper is as follows. In Section 2 we present a kernel estimator of the failure rate at x_0 , $h(n,x_0)$, introduce some notation and terminology, and discuss some general properties of $h(n,x_0)$.

In Section 3 we first prove a theorem (Theorem 3.1) which gives us the asymptotic bias, and enables us to obtain the asymptotic MSE of $h(n,x_0)$. An important result of Section 3 is a "saturation theorem" (Theorem 3.2). This theorem establishes the fact that if the kernel used to obtain $h(n,x_0)$ is nonnegative and absolutely integrable, then there is a limit beyond which the rate of convergence of the bias and the MSE does not increase, even if greater smoothness of the failure rate is assumed. We show that the best possible rate of convergence of the MSE using a nonnegative kernel is of the order $n^{-4/5}$ [Equation (3.8)]. Kernels which are not restricted to be nonnegative are next

introduced [Equation (3.9)]. The rate of convergence of the bias of $h(n,x_0)$ using such kernels is derived (Theorem 3.3). We conclude Section 3 by showing that the best possible rate of convergence of the MSE using kernels in A_m [cf. Equation (3.9)] is of the order $n^{-\frac{2m}{2m+1}}$.

In Section 4, we introduce the generalized jackknife procedure, and review briefly those properties of the procedure that are of interest to us. We show how the GJ can be used to construct the suitable kernels which are not restricted to be nonnegative. In particular, we show how the GJ when used once, can, under certain conditions, produce an estimator for which the best possible rate of convergence of the MSE is of the order $n^{-12/13}$. This is a clear improvement over the original kernel estimators whose rate of convergence of the MSE is of the order $n^{-4/5}$. We conclude Section 4 by giving some examples.

In Section 5 we concern ourselves with using the generalized jackknife procedure over and over again, an indefinite number of times. We show how this procedure enables us to produce estimators for which the rate of convergence of the MSE can be brought as close to n^{-1} as is desired. We give a few examples to illustrate the effect of a successive use of the GJ procedure. These examples motivate us to conjecture that the effect of an indefinite use of the GJ effectively leads us to kernels which are not absolutely integrable. This topic is explored in greater detail in another paper [Singpurwalla and Wong (1980)].

2. Kernel Estimates of the Failure Rate

Suppose that the time to failure of a device is a nonnegative random variable X with an absolutely continuous distribution function F and a probability density function f. The failure rate at point x_0, $h(x_0)$, for $F(x_0) \neq 1$, is defined as

$$h(x_0) = \frac{f(x_0)}{1 - F(x_0)} . \tag{2.1}$$

Given an ordered sample of n lifetimes $X_{(1)} \leq X_{(2)} \leq \cdots \leq X_{(n)}$ from F, our objective is to estimate $h(x_0)$ under some very general assumptions on F and f.

We shall consider the following kernel type estimators originally proposed by Watson and Leadbetter (1964a, 1964b), and considered more recently by Rice and Rosenblatt (1976) and Sethuraman and Singpurwalla (1978).

Definition 2.1: A kernel estimate, $h(n,x_0)$, of the failure rate $h(x)$ at the point x_0, is defined as

$$h(n,x_0) = \sum_{j=1}^{n} \frac{1}{b(n)} K\left(\frac{X_{(j)}-x_0}{b(n)}\right) \frac{1}{n-j+1} , \tag{2.2}$$

where $K(\cdot)$ is a Borel function called the *kernel* of the estimator, and $b(n)$ is a sequence of positive functions of n such that

$$\begin{aligned} &\text{(i)} \quad \lim_{n\to\infty} b(n) = 0 , \quad \text{and} \\ &\text{(ii)} \quad \lim_{n\to\infty} nb(n) = \infty . \end{aligned} \tag{2.3}$$

A motivation for these estimates is that if $X_{(1)} \leq X_{(2)} \leq \cdots \leq X_{(n)}$ come from an arbitrary distribution F, then the maximum likelihood estimate of f is a discrete distribution with a probability mass

of $1/n$ at $X_{(j)}$ [Grenander (1956)]; an estimate of the failure rate at $X_{(j)}$ is therefore $1/(n-j)$. In order to avoid the infinite at $j=n$, we change our estimate of the failure rate at $X_{(j)}$ to $1/(n-j+1)$, and smear this quantity out, continuously smoothing according to the kernel $K(\cdot)$.

We restrict our consideration to kernels which satisfy the following conditions:

$$\begin{cases} \text{(i)} & \sup_X |K(x)| < \infty , \\ \text{(ii)} & K(-x) = K(x) , \\ \text{(iii)} & \int K(x)dx = 1 ; \end{cases} \tag{2.4a}$$

$$\begin{cases} \text{(iv)} & \int |K(x)|dx < \infty , \\ \text{(v)} & \lim_{|X|\to\infty} |xK(x)| = 0 . \end{cases} \tag{2.4b}$$

When (iv) above holds, we shall say that $K \varepsilon L^1$.

Examples of some kernels which satisfy (2.4a) and (2.4b) are the following:

$$K(x) = \begin{cases} \frac{1}{2} & , \text{ if } |x| \leq 1 \\ 0 & , \text{ if } |x| > 1 \end{cases} \qquad \text{(rectangular)}$$

$$K(x) = \begin{cases} 1 - |x| , & \text{if } |x| \leq 1 \\ 0 & , \text{ if } |x| > 1 \end{cases} \qquad \text{(triangular)}$$

$$K(x) = (2\pi)^{-1/2} e^{-x^2/2} \qquad \text{(Weierstrass)}$$

$$K(x) = \frac{1}{2} e^{-|x|} \qquad \text{(Picard)}$$

$$K(x) = \pi^{-1} \left(1+x^2\right)^{-1} \qquad \text{(Cauchy)}$$

$$K(x) = \frac{1}{2\pi}\left(\frac{\sin \frac{x}{2}}{\frac{x}{2}}\right)^2 .$$

We remark that all the kernels given above are nonnegative.

We shall find it convenient to introduce the following definitions.

Definition 2.2: A sequence of functions $\{\delta_n(x)\}$ is called a *δ-function sequence* if $\delta_n(x)$ can be written in the form

$$\delta_n(x) = \frac{1}{b(n)} K\left(\frac{x}{b(n)}\right) ,$$

where $K(\cdot)$ is a kernel which satisfies (2.4a) and (2.4b), and $b(n)$ is a sequence of nonnegative decreasing functions which satisfy (2.3).

Thus (2.2) can be written as

$$h(n,x_0) = \sum_{j=1}^{n} \frac{\delta_n\left(X_{(j)}-x_0\right)}{n-j+1} . \tag{2.5}$$

Definition 2.3: For a given δ-function sequence $\delta_n(\cdot)$, a distribution function F is said to be in the *class* C_δ , if for any fixed x_0 , and for any fixed $\lambda > 0$, there exists a $G_\lambda > 0$ such that

$$\frac{|\delta_n(x-x_0)|}{1 - F(x)} \leq G_\lambda , \tag{2.6}$$

for all $|x-x_0| \geq \lambda$ and for all sufficiently large n .

Note that if the kernel K is *bandlimited*, that is, if $K(x) = 0$ for all $|x| > C$, for some finite real value $C > 0$, then C_δ is the class of all distribution functions.

Before describing the properties of the estimator $h(n,x_0)$, we should indicate a few abbreviations. If a_n and b_n are two sequences, then "$a_n \sim b_n$" is read a_n is asymptotically equivalent to b_n , and means that the ratio of a_n to b_n has limit one. The notation

"$a_n = o(b_n)$" means that the ratio of a_n to b_n has limit 0, and "$a_n = O(b_n)$" means that the absolute value of the ratio is bounded in the limit. The terms $o(b_n)$ and $O(b_n)$ are frequently used to represent some unknown function of n which has the appropriate property.

The following theorem establishes the asymptotic unbiasedness and the consistency of the estimator (2.2).

Theorem 2.1 [Watson and Leadbetter (1964a)]: Let $\{\delta_n(x)\}$ be a δ-function sequence and let $F(x)$ be a distribution function in C_δ. If $h(x)$ is continuous at x_0, and if $F(x_0) < 1$, then $h(n,x_0)$ given by (2.2) is an asymptotically unbiased estimator of $h(x_0)$.

Furthermore, if $\alpha_n = \int \delta_n^2(x)dx < \infty$ and $\alpha_n = o(n)$, then $h(n,x_0)$ is consistent with an asymptotic variance

$$\mathrm{Var}[h(n,x_0)] \sim \frac{\alpha_n}{n} \frac{h(x_0)}{1-F(x_0)} . \tag{2.7}$$

That is, the variance of $h(n,x_0)$ goes to zero at the same rate at which α_n/n goes to zero.

3. Rates of Convergence of the Bias and the Mean Square Errors

In order to compare estimates of $h(x_0)$ using different kernels, we will have to study the rates of convergence of their bias and their mean square errors. In general, the rate of decrease of the bias and the MSE depends on the particular kernel (or the δ-function sequence) that is chosen. Furthermore, for a given kernel, the rate of convergence improves with the smoothness of the failure rate function h. However, for some kernels (in particular, the nonnegative kernels), there may be a limit beyond which the rates of convergence do not increase, even if greater smoothness of h is assumed. This phenomenon is called "saturation" [see Shapiro (1969)].

To see this, we shall first give the following theorem on the asymptotic bias of $h(n,x_0)$.

Theorem 3.1: Let $X_1, X_2, \ldots, X_n$ be a random sample of lifetimes from an absolutely continuous distribution function F, and probability density function f; x_0 is a continuity point of the failure rate h of F and $F(x_0) < 1$. Let $h(n,x_0)$, an estimate of $h(x_0)$, be given by (2.5). If $F \varepsilon C_\delta$, and if, for some positive integer $m \geq 1$, $h \varepsilon C^m$ (i.e., h is m times continuously differentiable), $\lim_{n\to\infty} nb^m(n) = \infty$, and the kernel K is such that $x^m K \varepsilon L^1$, then the bias of $h(n,x_0)$ is given by

$$\text{Bias}[h(n,x_0)] = \sum_{j=1}^{m} b^j(n) \frac{h^{(j)}(x_0)}{j!} \int x^j K(x)dx + o\left(b^m(n)\right) . \quad (3.1)$$

<u>Proof:</u>

$$E[h(n,x_0)] = E\left\{\sum_{j=1}^{n} \frac{\delta_n(X_{(j)}-x_0)}{n-j+1}\right\}$$

$$= \sum_{j=1}^{n} \int \frac{1}{n-j+1}\, \delta_n(u-x_0)\, f_{X_{(j)}}(u)du$$

$$= \sum_{j=1}^{n} \int \frac{1}{n-j+1}\, \delta_n(u-x_0)\, \frac{n!}{(j-1)!(n-j)!}\, (F(u))^{j-1} f(u)(1-F(u))^{n-j} du$$

$$= \int \sum_{j=1}^{n} \binom{n}{j-1}\, \delta_n(u-x_0) h(u)(F(u))^{j-1}(1-F(u))^{n-j+1} du$$

$$= \int h(u)\, \delta_n(u-x_0)\left\{\sum_{j=1}^{n+1} \binom{n}{j-1}\, (F(u))^{j-1}(1-F(u))^{n-j+1} - F^n(u)\right\} du$$

$$= \int h(u)\, \delta_n(u-x_0)\, [1-F^n(u)]du$$

$$= \int h(u)\, \delta_n(u-x_0)du - \int h(u)\, \delta_n(u-x_0)F^n(u)du\ . \tag{3.2}$$

Consider the first term of the above expression; note that it can also be written as

$$\int h(u)\, \frac{1}{b(n)}\, K\left(\frac{u-x_0}{b(n)}\right)du = \int h\big(x_0+vb(n)\big)\, K(v)dv\ .$$

Using a Taylor's series expansion about x_0, we can write the first term of (3.2) as

$$\int\left[h(x_0) + \sum_{j=1}^{m} \frac{b^j(n)}{j!}\, h^{(j)}(x_0)v^j + o\big(b^m(n)\big)\right] K(v)dv$$

$$= h(x_0) + \sum_{j=1}^{m} \frac{b^j(n)}{j!}\, h^{(j)}(x_0) \int v^j K(v)dv + o\big(b^m(n)\big)\ ,$$

since $h \in C^m$ and $x^m K \in L^1$.

In order to see that the second term of (3.2) is $o(b^m(n))$, we choose a $\lambda > 0$ so that $F(x_0+\lambda) < 1$, and $h(u)$ is bounded in $|u-x_0| \leqq \lambda$. Then

$$\left|\int h(u)\delta_n(u-x_0)F^n(u)du\right| \leqq \int h(u) \left|\delta_n(u-x_0)\right| F^n(u)du$$

$$= \int_{|u-x_0|>\lambda} f(u) \frac{|\delta_n(u-x_0)|}{1-F(u)} F^n(u)du + \int_{|u-x_0|\leqq\lambda} h(u) \left|\delta_n(u-x_0)\right| F^n(u)du$$

$$\leqq G_\lambda \int_0^1 F^n dF + \text{const} \cdot F^n(x_0+\lambda) \qquad [\text{by } (2.6)]$$

$$= \frac{G_\lambda}{n+1} + \text{const} \cdot F^n(x_0+\lambda) .$$

Using the above inequality we observe that

$$\frac{1}{b^m(n)} \left|\int h(u)\delta_n(u-x_0)F^n(u)du\right| \leqq \frac{G_\lambda}{(n+1)b^m(n)} + \text{const} \cdot \frac{F^n(x_0+\lambda)}{b^m(n)} \to 0 ,$$

as $n \to \infty$; thus

$$\int h(u)\delta_n(u-x_0)F^n(u)du = o(b^m(n)) .$$

The statement of the theorem follows if we combine our results on the two terms of (3.2). //

Since K is assumed to be an even function [i.e., $K(-x) = K(x)$], the bias can be written (more precisely) as

$$\text{Bias}[h(n,x_0)] = \sum_{k=1}^{t} b^{2k}(n) \frac{h^{(2k)}(x_0)}{(2k)!} \int x^{2k}K(x)dx + o(b^{2t}(m)) , \quad (3.3)$$

where $m = 2t$ for some positive integer t .

Note that Theorem 3.1 is valid for all kernels K, nonnegative or otherwise, which satisfy the conditions of the theorem. In order to discuss a saturation result for $h(n,x_0)$ when K is nonnegative, we look at a special case of (3.1), when $m=2$; that is, when

(i) $x^2 K \in L^1$,

(ii) $\lim_{n\to\infty} nb^2(n) = \infty$, and (3.4)

(iii) $h \in C^2$,

we have as a corollary to Theorem 3.1

<u>Theorem 3.2</u> (<u>pointwise saturation theorem</u>): Let $h(n,x_0)$ be the failure rate estimation of $h(x_0)$ given by (2.5); x_0 is a continuity point of the failure rate h and $F(x_0) < 1$. Under conditions (3.4) and if $F \in C_\delta$, then

$$\lim_{n\to\infty} \frac{\text{Bias}[h(n,x_0)]}{b^2(n)} = \frac{h''(x_0)}{2!} \int x^2 K(x)dx \ . \tag{3.5}$$

<u>Proof</u>: Follows from 3.1 and the fact that K is even $\int xK(x) = 0$. //

For those kernels K which are nonnegative and satisfy the conditions of Theorem 3.2 (e.g., when K is the rectangular, triangular, Weierstrass, or Picard), $\int x^2 K(x)$ is *always* nonzero. If $h''(x_0) \neq 0$, then the rate of convergence of the bias, according to (3.5), is no faster than $b^2(n)$. Thus we conclude that the best possible rate of decrease of the bias of $h(n,x_0)$ with nonnegative kernels is $b^2(n)$.

We now consider the mean square error of the estimator $h(n,x_0)$ under the assumptions of Theorem 3.2. Since

$$\text{Bias}[h(n,x_0)] \sim b^2(n) \frac{h''(x_0)}{2!} \int x^2 K(x)dx \ ,$$

and the asymptotic variance of $h(n,x_0)$ given by (2.7) is

$$\text{Var}[h(n,x_0)] \sim \frac{\int \delta_n^2(x)dx}{n} \frac{h(x_0)}{1-F(x_0)} \ ,$$

the asymptotic MSE of $h(n,x_0)$ is

$$\mathrm{MSE}[h(n,x_0)] = E\left[\left(h(n,x_0) - h(x_0)\right)^2\right] = \mathrm{Var}[h(n,x_0)] + \mathrm{Bias}^2[h(n,x_0)]$$

$$\sim \frac{\int \delta_n^2(x)dx}{n} \frac{h(x_0)}{1-F(x_0)} + \left[b^2(n) \frac{h''(x_0)}{2!} \int x^2 K(x)dx\right]^2 ,$$

or

$$\mathrm{MSE}[h(n,x_0)] \sim \frac{1}{nb(n)} \frac{h(x_0)}{1-F(x_0)} \int K^2(v)dv + \left[b^2(n) \frac{h''(x_0)}{2!} \int x^2 K(x)dx\right]^2 . \quad (3.6)$$

Given h and K , we can, for a fixed value of n , find that $b(n)$ which minimizes the asymptotic MSE of $h(n,x_0)$ by solving

$$2\left\{2b^3(n)\left[\frac{h''(x_0)}{2!}\int x^2K(x)dx\right]^2\right\} = \frac{1}{nb^2(n)} \frac{h(x_0)}{1-F(x_0)} \int K^2(v)dv .$$

We obtain, if $h''(x_0) \neq 0$ and $\int x^2 K(x)dx \neq 0$,

$$b(n) = \left\{\frac{\frac{h(x_0)}{1-F(x_0)} \int K^2(v)dv}{2\cdot 2\left[h''(x_0)\frac{\int x^2K(x)dx}{2!}\right]^2}\right\}^{1/5} n^{-1/5} . \quad (3.7)$$

For this value of $b(n)$, we have the optimum value of the MSE,

$$\mathrm{MSE}[h(n,x_0)]_{\mathrm{opt}} \sim (2\cdot 2+1)\left[\frac{\int K^2(v)dv}{2\cdot 2} \frac{h(x_0)}{1-F(x_0)}\right]^{4/5}\left[\frac{\int x^2K(x)dx}{2!} h''(x_0)\right]^{2/5} n^{-4/5} . \quad (3.8)$$

Based upon (3.7) and (3.8) we conclude that if $b(n) = O(n^{-1/5})$, then $\mathrm{MSE}[h(n,x_0)]_{\mathrm{opt}} \sim O(n^{-4/5})$. Thus if the kernel K is restricted to be *nonnegative* and if K is an L^1 kernel (i.e., $K \in L^1$) , then the best possible rate of convergence of the MSE of $h(n,x_0)$ is $O(n^{-4/5})$, regardless of the smoothness of h .

The above results motivate us to consider kernels which are not restricted to be nonnegative, and which are not L^1. In Singpurwalla and Wong (1980) we consider a kernel which is not L^1 and which can take negative values. However, we shall first show here (Theorem 3.3) that we can obtain an improvement in the rate of convergence of the MSE of $h(n,x_0)$ when K is not restricted to be nonnegative, but is still an L^1 kernel; although one might think that nonnegative kernels might provide the best estimates since the failure rates are nonnegative. It is to be expected that the price that we will have to pay for obtaining faster rates of convergence of the MSE using such kernels is that the resulting estimate of the failure rate could be negative at some points.

Let A_m (where $m \geq 2$ is a positive integer) be the class of all real valued Borel measurable bounded functions K (kernels), which satisfy conditions (2.4) and the following condition:

$$\int x^j K(x)dx = 0 , \quad \text{for } j=1,2,\ldots,m-1 . \tag{3.9}$$

All kernels (nonnegative or otherwise) which satisfy (2.4) also satisfy (3.9) for $m=2$; thus A_2 is the class of all kernels which satisfy (2.4). For $m \geq 3$, the class A_m contains no nonnegative functions and its elements will therefore lead to possibly negative failure rate estimates if $K \varepsilon A_m$ is used in estimation. Since K is an even function, $K \varepsilon A_{2k-1}$ implies that $K \varepsilon A_{2k}$, for any $k=2,3,\ldots$; thus the class A_m need only be defined for an even integer $m > 1$.

<u>Theorem 3.3</u>: Let $X_1, X_2, \ldots, X_n$ be a random sample of lifetimes from an absolutely continuous distribution function F, and probability density function f ; x_0 is a continuity point of the failure rate h of F and $F(x_0) > 1$. Let $h(n,x_0)$ be an estimate of $h(x_0)$, given by (2.5). If $F \varepsilon C_\delta$, and if, for some positive integer $m \geq 1$, the following conditions hold:

$$\begin{cases} \text{(i)} & K \in A_m , \\ \text{(ii)} & x^m K \in L^1 , \\ \text{(iii)} & \int x^m K(x) \neq 0 , \end{cases} \tag{3.10a}$$

$$\begin{cases} \text{(iv)} & \lim_{n\to\infty} nb^m(n) = \infty , \\ \text{(v)} & h \in C^m ; \end{cases} \tag{3.10b}$$

then

$$\lim_{n\to\infty} \frac{1}{b^m(n)} \text{Bias}[h(n,x_0)] = \frac{h^{(m)}(x_0)}{m!} \int x^m K(x)dx .$$

Proof: Analogous to that of Theorem 3.2, but in the light of conditions (3.10). //

Thus, the bias of the estimator $h(n,x_0)$ based on a kernel K which satisfies conditions (3.10a) cannot decrease any faster than $b^m(n)$; in fact,

$$\text{Bias}[h(n,x_0)] \sim b^m(n) \frac{h^m(x_0)}{m!} \int x^m K(x)dx . \tag{3.11}$$

Since the asymptotic variance of $h(n,x_0)$ is still of the form

$$\text{Var}[h(n,x_0)] \sim \frac{\int \delta_n^2(x)dx}{n} \frac{h(x_0)}{1-F(x_0)} ,$$

the asymptotic MSE of $h(n,x_0)$ with $K \in A_m$ is

$$\text{MSE}[h(n,x_0)] \sim \frac{1}{nb(n)} \frac{h(x_0)}{1-F(x_0)} \int K^2(v)dv + \left[b^m(n) \frac{h^m(x_0)}{m!} \int x^m K(x)dx \right]^2 . \tag{3.12}$$

Given h and K , we can, for a fixed value of n , find that $b(n)$ which minimizes the asymptotic MSE of $h(n,x_0)$. This value is

$$b(n) = \left\{ \frac{\frac{h(x_0)}{1-F(x_0)} \int K^2(v)dv}{2m\left[h^m(x_0) \frac{\int x^m K(x)dx}{m!}\right]^2} \right\}^{\frac{1}{2m+1}} n^{-\frac{1}{2m+1}} . \qquad (3.13)$$

For this value of $b(n)$, the optimal value of the MSE is

$$MSE[h(n,x_0)]_{opt}$$

$$\sim (2m+1)\left[\frac{\int K^2(v)dv}{2m} \frac{h(x_0)}{1-F(x_0)}\right]^{\frac{2m}{2m+1}} \left[\frac{\int x^m K(x)dx}{m!} h^m(x_0)\right]^{\frac{2}{2m+1}} n^{-\frac{2m}{2m+1}} . \qquad (3.14)$$

Based upon the above results we state that if $b(n) = O(n^{-\frac{1}{2m+1}})$, then $MSE[h(n,x_0)]_{opt} = O(n^{-\frac{2m}{2m+1}})$; i.e., $MSE[h(n,x_0)]_{opt} \to 0$ as $n^{-\frac{2m}{2m+1}}$. We contrast this result to our previous result using nonnegative kernels for which $MSE[h(n,x_0)] \to 0$ as $n^{-4/5}$, which is slower than $n^{-\frac{2m}{2m+1}}$ for $m \geq 3$ when using kernels which are not restricted to be nonnegative. We have proved the following important result.

Theorem 3.4: Under the conditions of Theorem 3.3, the asymptotically optimal rate of convergence of the MSE of the failure rate estimator $h(n,x_0)$ is of the order $n^{-\frac{2m}{2m+1}}$.

4. The Generalized Jackknife and Kernels in A_m

Finding kernels $K \in A_2$ is quite straightforward, since many of the continuous density functions which are symmetric about 0 satisfy the conditions of Theorem 3.2. The best possible rate of convergence of the MSE of $h(n,x_0)$ based on these kernels is $n^{-4/5}$. However, as we have seen before, we can obtain estimators with a faster rate of convergence in bias and MSE by using kernels in A_m, $m \geq 3$. In this event, the bias decreases in the order of $b^m(n)$ and the optimal MSE tends to zero as $n^{-\frac{2m}{2m+1}}$. Our next objective is to discuss a procedure which leads us to such kernels.

In this section we shall discuss the generalized jackknife method [Schucany, Gray, and Owen (1971)] of combining estimators, and show that this procedure, which is typically used to remove the bias and reduce the MSE, essentially leads us to kernels K which belong to A_m, $m \geq 3$.

We shall first give a brief introduction to the generalized jackknife method of combining estimators.

<u>Definition 4.1</u> [Gray and Schucany (1972)]: Let $\hat{\theta}_1$ and $\hat{\theta}_2$ be two estimators of θ. Then for any real number $R \neq 1$, we define the *generalized jackknife* $\tilde{\theta}$ of θ by

$$\tilde{\theta} = \frac{\hat{\theta}_1 - R\hat{\theta}_2}{1 - R}.$$

A trivial, though important, property of the estimator $\tilde{\theta}$ is given by the following theorem.

<u>Theorem 4.1</u> [Gray and Schucany (1972)]: If $E(\hat{\theta}_j) = \theta + b_j(n,\theta)$, $j=1,2$, $b_2(n,\theta) \neq 0$, and $R = \dfrac{b_1(n,\theta)}{b_2(n,\theta)} \neq 1$, then $\tilde{\theta}$ is an unbiased estimator of θ.

In particular, if the biases of $\hat{\theta}_1$ and $\hat{\theta}_2$ have the forms $b_j(n,\theta) = f_j(n)b(\theta)$, $j=1,2$, then, under the conditions of Theorem 4.1, the estimator $\tilde{\theta}$ is of the form

$$\tilde{\theta} = \frac{\hat{\theta}_1 - \dfrac{f_1(n)}{f_2(n)}\hat{\theta}_2}{1 - \dfrac{f_1(n)}{f_2(n)}} = \frac{f_2(n)\hat{\theta}_1 - f_1(n)\hat{\theta}_2}{f_2(n) - f_1(n)} = \frac{\det\begin{bmatrix} \hat{\theta}_1 & \hat{\theta}_2 \\ f_1(n) & f_2(n) \end{bmatrix}}{\det\begin{bmatrix} 1 & 1 \\ f_1(n) & f_2(n) \end{bmatrix}} ,$$

and $E(\tilde{\theta}) = \theta$.

The idea of Definition 4.1 and Theorem 4.1 can be generalized to include three or more estimators.

<u>Definition 4.2</u>: Let $\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_{k+1}$ be $k+1$ estimators of θ based on the same random sample $X_1, X_2, \ldots, X_n$. Further, let a_{ij} , $i=1,2,\ldots,k$; $j=1,2,\ldots,k+1$ be real numbers such that

$$\det\begin{bmatrix} 1 & 1 & \cdots & 1 \\ a_{11} & a_{12} & \cdots & a_{1,k+1} \\ \vdots & \vdots & & \vdots \\ a_{k1} & a_{k2} & \cdots & a_{k,k+1} \end{bmatrix} = \Delta \neq 0 .$$

Thus the *generalized jackknife* estimator is defined as

$$\tilde{\theta} = \frac{\det\begin{bmatrix} \hat{\theta}_1 & \hat{\theta}_2 & \cdots & \hat{\theta}_{k+1} \\ a_{11} & a_{12} & \cdots & a_{1,k+1} \\ \vdots & \vdots & & \vdots \\ a_{k1} & a_{k2} & \cdots & a_{k,k+1} \end{bmatrix}}{\Delta} .$$

When the bias in the estimators $\hat{\theta}_1, \ldots, \hat{\theta}_{k+1}$ can be written as the product of a function of n and a function of θ , we have the following result.

Underlined Theorem 4.2: If $E(\hat{\theta}_j) = \theta + \sum_{i=1}^{\infty} f_{ij}(n)b_i(\theta)$, $j=1,2,\ldots,k+1$, and if

$$\det \begin{bmatrix} 1 & 1 & \cdots & 1 \\ f_{11}(n) & f_{12}(n) & \cdots & f_{1,k+1}(n) \\ \vdots & \vdots & & \vdots \\ f_{k1}(n) & f_{k2}(n) & \cdots & f_{k,k+1}(n) \end{bmatrix} = \Delta \neq 0 ,$$

then $E(\tilde{\theta}) = \theta + B_G(n,\theta)$, where

$$B_G(n,\theta) = \frac{\det \begin{bmatrix} B_1 & B_2 & \cdots & B_{k+1} \\ f_{11}(n) & f_{12}(n) & \cdots & f_{1,k+1}(n) \\ \vdots & \vdots & & \vdots \\ f_{k1}(n) & f_{k2}(n) & \cdots & f_{k,k+1}(n) \end{bmatrix}}{\Delta}$$

and $B_j = \sum_{i=k+1}^{\infty} f_{ij}(n)b_i(\theta)$, $j=1,2,\ldots,k+1$.

Corollary 4.3: If $E(\hat{\theta}_j) = \theta + \sum_{i=1}^{k} f_{ij}(n)b_i(\theta)$, $j=1,2,\ldots,k+1$, then $E(\tilde{\theta}) = \theta$.

For a proof of the above results, we refer the reader to Schucany, Gray, and Owen (1971) or Gray and Schucany (1972). Miller (1974, 1978) has given an up to date summary of the jackknife and its various applications.

It is apparent from Theorem 4.1 and Corollary 4.3 that generalized jackknifing is a way of bias reduction. It is with this thought in mind that we consider combinations of estimators $h(n,x_0)$ based upon different kernels in A_2 (using the generalized jackknife) to arrive at estimators

of $h(x_0)$ which have a smaller bias. It turns out, as we shall soon see, that the generalized jackknifed estimator of $h(x_0)$ is precisely that which we would have obtained by considering a kernel in A_m, $m \geq 3$. Thus jackknifing kernel estimates of the failure rate based upon kernels in A_2 will produce estimators which have faster rates of convergence of the bias and the mean square error, but which by virtue of the fact that they could also have been produced by kernels in A_m, $m \geq 3$, may be negative.

Let us consider two estimators of $h(x_0)$ based on kernels K_1 and K_2, where K_1 and K_2 belong to A_2 (not necessarily to A_m, $m \geq 3$); thus we have

$$h_i(n,x_0) = \frac{1}{b_i(n)} \sum_{j=1}^{n} \frac{1}{n-j+1} K_i\left(\frac{X_{(j)}-x_0}{b_i(n)}\right), \quad i=1,2 . \tag{4.1}$$

The generalized jackknife estimator $\tilde{h}(n,x_0)$ of h_1 and h_2 is

$$\tilde{h}(n,x_0) = \frac{h_1(n,x_0) - Rh_2(n,x_0)}{1 - R},$$

where $R \neq 1$ is a constant to be determined. If the estimators h_1 and h_2 are such that the conditions of Theorem 3.1 hold for $m = 2t \geq 6$, then from (3.3)

$$
\begin{aligned}
E_X[\tilde{h}(n,x_0) - h(x_0)] &= E\left[\frac{h_1(n,x_0) - Rh_2(n,x_0)}{1 - R} - h(x_0)\right] \\
&= \frac{1}{1-R}\left\{E[h_1(n,x_0) - h(x_0)] - RE[h_2(n,x_0) - h(x_0)]\right\} \\
&= \frac{1}{1-R}\left\{\sum_{k=1}^{t} \frac{h^{(2k)}(x_0)}{(2k)!} b_1^{2k}(n) I(K_1,2k)\right. \\
&\qquad \left. - R\sum_{k=1}^{t} \frac{h^{(2k)}(x_0)}{(2k)!} b_2^{2k}(n) I(K_2,2k)\right\} \\
&\qquad + o\left(b_1^{2t}(n)\right) + o\left(b_2^{2t}(n)\right) \\
&= \frac{1}{1-R}\sum_{k=1}^{t}\left[b_1^{2k}(n) I(K_1,2k) - Rb_2^{2k}(n) I(K_2,2k)\right] \frac{h^{(2k)}(x_0)}{(2k)!} \\
&\qquad + o\left(b_1^{2t}(n)\right) + o\left(b_2^{2t}(n)\right) \qquad (4.3)
\end{aligned}
$$

where $I(K,q) = \int x^q K(x)dx$. If we set

$$
R = \frac{b_1^2(n) I(K_1,2)}{b_2^2(n) I(K_2,2)} , \qquad (4.4)
$$

then $b_1^2(n)I(K_1,2) - Rb_2^2(n)I(K_2,2) = 0$, and the leading bias term of $\tilde{h}(n,x_0)$, that is, the term containing $h''(x_0)$ in (4.3), is eliminated. The estimator $h(n,x_0)$ now becomes

$\tilde{h}(n,x_0)$

$$= \frac{\dfrac{1}{b_1(n)} \displaystyle\sum_{j=1}^{n} \frac{1}{n-j+1} K_1\left(\frac{X_{(j)}-x_0}{b_1(n)}\right) - \dfrac{b_1^2(n) I(K_1,2)}{b_2^2(n) I(K_2,2)} \dfrac{1}{b_2(n)} \displaystyle\sum_{j=1}^{n} \frac{1}{n-j+1} K_2\left(\frac{X_{(j)}-x_0}{b_2(n)}\right)}{1 - \dfrac{b_1^2(n) I(K_1,2)}{b_2^2(n) I(K_2,2)}}$$

$$= \frac{1}{b_1(n)} \sum_{j=1}^{n} \frac{1}{n-j+1} \left[\frac{K_1\left(\dfrac{X_{(j)}-x_0}{b_1(n)}\right) - \dfrac{b_1^3(n)}{b_2^3(n)} \dfrac{I(K_1,2)}{I(K_2,2)} K_2\left(\dfrac{X_{(j)}-x_0}{b_2(n)}\right)}{1 - \dfrac{b_1^2(n)}{b_2^2(n)} \dfrac{I(K_1,2)}{I(K_2,2)}}\right]$$

$$= \frac{1}{b_1(n)} \sum_{j=1}^{n} \frac{1}{n-j+1} \tilde{K}\left(\frac{X_{(j)}-x_0}{b_1(n)}\right) , \tag{4.5}$$

where

$$\tilde{K}(u) = \frac{K_1(u) - c^3(u) \dfrac{I(K_1,2)}{I(K_2,2)} K_2\big(c(n)u\big)}{1 - c^2(n) \dfrac{I(K_1,2)}{I(K_2,2)}} , \tag{4.6}$$

and

$$c(n) = \frac{b_1(n)}{b_2(n)} .$$

We can write $\tilde{K}(u) = \alpha K_1(u) - \beta K_2\big(c(n)u\big)$, where

$$\alpha = \frac{1}{1 - c^2(n) \dfrac{I(K_1,2)}{I(K_2,2)}} , \quad \text{and} \quad \beta = \frac{c^3(n) \dfrac{I(K_1,2)}{I(K_2,2)}}{1 - c^2(n) \dfrac{I(K_1,2)}{I(K_2,2)}} . \tag{4.7}$$

Thus, the generalized jackknife which enables us to combine estimators $h_i(n,x_0)$ based upon kernels K_i, $i=1,2$, which belong to A_2 (and not necessarily to A_m, $m \geq 3$), and by choosing R according to (4.4), we have produced an estimator $\tilde{h}(n,x_0)$ based upon a kernel $\tilde{K}$.

We claim that $\tilde{K} \varepsilon A_4$. To see this, we first note that

$$I(\tilde{K},0) = \int \tilde{K}(u)du = \alpha \int K_1(u)du - \beta \int K_2(c(n)u)du$$

$$= \alpha - \frac{\beta}{c(n)} = 1 .$$

We can now easily verify that $\tilde{K}$ satisfies (2.4). To see that $\tilde{K}$ satisfies (3.9) for $m=4$, it suffices to show that $I(\tilde{K},1) = I(\tilde{K},2) = I(\tilde{K},3) = 0$. Since K_1 and K_2 are even functions, $I(\tilde{K},1) = I(\tilde{K},3) = 0$. Now,

$$I(\tilde{K},2) = \int u^2\tilde{K}(u)du = \int \alpha u^2 K_1(u)du - \int \beta u^2 K_2(c(n)u)du$$

$$= \alpha I(K_1,2) - \frac{\beta}{c^3(n)} I(K_2,2) = 0 . \qquad //$$

The optimal MSE of $\tilde{h}(n,x_0)$ with kernel $\tilde{K}$ is given by (3.14), with $m=4$, as

$$MSE[\tilde{h}(n,x_0)]_{opt}$$

$$\sim (2\cdot 4+1)\left[\frac{\int \tilde{K}^2(u)du}{2\cdot 4}\,\frac{h(x_0)}{1-F(x_0)}\right]^{8/9}\left[\frac{\int u^4\tilde{K}(u)du}{4!}\,h^{(4)}(x_0)\right]^{2/9} n^{-8/9} .$$

Note that by choosing R according to (4.4), we have eliminated the leading bias term of (4.3). We are still at liberty to choose $b_1(n)$ and $b_2(n)$ in any manner, provided $R \neq 1$. Clearly, by choosing $b_1(n)$ and $b_2(n)$ in such a manner that

$$b_1^4(n)I(K_1,4) - Rb_2^2(n)I(K_2,4) = 0 ,$$

that is

$$\frac{b_1^2(n)}{b_2^2(n)} = \frac{I(K_2,4)I(K_1,2)}{I(K_1,4)I(K_2,2)} , \tag{4.8}$$

we can eliminate the second bias term in (4.3), that is, the term containing $h^{(4)}(x_0)$.

When (4.4) and (4.8) are used, then the kernel $\tilde{K}$ given in (4.6) belongs to A_6 , since

$$\begin{aligned}
I(\tilde{K},4) &= \int u^4\tilde{K}(u)du = \alpha \int u^4 K_1(u)du - \beta \int u^4 K_2\big(c(n)u\big)du \\
&= \alpha I(K_1,4) - \frac{\beta}{c^5(n)} I(K_2,4) \\
&= \frac{I(K_1,4) - \frac{I(K_1,2)}{I(K_2,2)} c^3(n) \frac{1}{c^5(n)} I(K_2,4)}{1 - \frac{I(K_1,2)}{I(K_2,2)} c^2(n)} \\
&= \frac{I(K_1,4) - \frac{I(K_1,2)}{I(K_2,2)} \frac{I(K_1,4)I(K_2,2)}{I(K_2,4)I(K_1,2)} I(K_2,4)}{1 - \frac{I(K_1,2)}{I(K_2,2)} c^2(n)} = 0 .
\end{aligned}$$

When $\tilde{K} \varepsilon A_6$, the asymptotic MSE of $\tilde{h}(n,x_0)$ is by (3.12),

$$MSE[\tilde{h}(n,x_0)] \sim \frac{1}{n\tilde{b}(n)} \frac{h(x_0)}{1-F(x_0)} \int \tilde{K}^2(u)du + \left[\frac{\tilde{b}^6(n)}{6!} h^{(6)}(x_0) \int u^6\tilde{K}(u)du\right]^2 .$$

Given h and $\tilde{K} \varepsilon A_6$, we can, for a fixed value of n , find that $b(n)$ which minimizes the asymptotic MSE of $\tilde{h}(n,x_0)$. This value is

$$\tilde{b}(n) = \left\{ \frac{\frac{h(x_0)}{1-F(x_0)} \int \tilde{K}^2(u)du}{2\cdot 6\left[h^{(6)}(x_0)\ \frac{\int u^6\tilde{K}(u)du}{6!}\right]^2} \right\}^{1/13} n^{-1/13} .$$

The optimal value of the MSE for this choice of $b(n)$ is

$$MSE[\tilde{h}(n,x_0)]_{opt}$$

$$\sim (2\cdot 6+1) \left[\frac{\int \tilde{K}(u)du}{2\cdot 6}\ \frac{h(x_0)}{1-F(x_0)}\right]^{12/13} \left[\frac{\int u^6\tilde{K}(u)du}{6!}\ h^{(6)}(x_0)\right]^{2/13} n^{-12/13} .$$

Since $MSE[\tilde{h}(n,x_0)]_{opt} \to 0$ as $n^{-12/13}$, the estimator $\tilde{h}(n,x_0)$ has a faster rate of convergence of the MSE than the original estimators $h_1(n,x_0)$ and $h_2(n,x_0)$, whose MSE's $\to 0$ as $n^{-4/5}$.

Thus we have seen how we can form linear combinations of two estimators, h_1 and h_2, based upon kernels in A_2 to obtain a new estimator $\tilde{h}$ with kernel $\tilde{K} \in A_6$. The new estimator has a smaller bias than the estimators h_1 and h_2, and has a faster rate of convergence of the MSE.

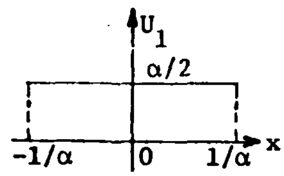
We close this section by noting that the higher order jackknife may be used to obtain linear combinations of several kernel estimators, if this is desired.

4.1 Examples.

We shall illustrate the material discussed in this section by considering some specific kernels in A_2, and illustrate the nature of the new kernels (in A_4 or A_6) which are obtained by jackknifing and an appropriate choice of other constants.

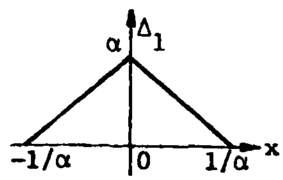
Example 4.1: Consider two estimators of $h(x_0)$, $h_i(n,x_0)$, $i=1,2$, based on the uniform kernel U_1 and the triangular kernel Δ_1, respectively. That is,

$$U_1(x) = \begin{cases} \frac{\alpha}{2}, & |x| \leq \frac{1}{\alpha} \\ 0, & |x| > \frac{1}{\alpha} \end{cases},$$



and

$$\Delta_1(x) = \begin{cases} \alpha(1-\alpha|x|), & |x| \leq \frac{1}{\alpha} \\ 0, & |x| > \frac{1}{\alpha} \end{cases}.$$



Now

$$I(U_1,2) = \int x^2 U_1(x)dx = \frac{1}{3\alpha^2},$$

$$I(U_1,4) = \int x^4 U_1(x)dx = \frac{1}{5\alpha^2},$$

$$I(\Delta_1,2) = \int x^2 \Delta_1(x)dx = \frac{1}{6\alpha^2}, \text{ and}$$

$$I(\Delta_1,4) = \int x^4 \Delta_1(x)dx = \frac{1}{15\alpha^2}.$$

If we choose $b_1(n)$, $b_2(n)$, and R in such a manner that

$$c^2(n) = \frac{b_1^2(n)}{b_2^2(n)} = \frac{I(\Delta_1,4)I(U_1,2)}{I(U_1,4)I(\Delta_1,2)} = \frac{2}{3}$$

and

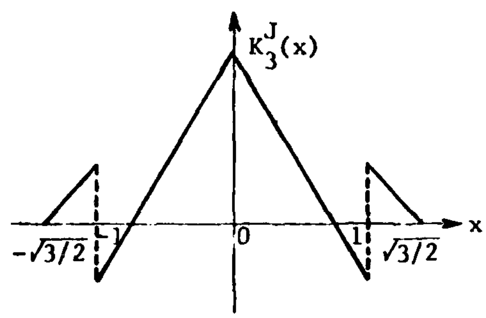
$$R = \frac{b_1^2(n)I(U_1,2)}{b_2^2(n)I(\Delta_1,2)} = \frac{4}{3},$$

then the new kernel K_3^J, given by (4.6), is

$$K_3^J(x) = \frac{U_1(x) - c^3(n)\,\frac{I(U_1,2)}{I(\Delta_1,2)}\,\Delta_1(c(n)x)}{1 - c^2(n)\,\frac{I(U_1,2)}{I(\Delta_1,2)}} = -3U_1(x) + 4\sqrt{\tfrac{2}{3}}\,\Delta_1\left(\sqrt{\tfrac{2}{3}}\,x\right) .$$

Note that $K_3^J \varepsilon A_6$. In particular, if we choose $\alpha = 1$ (in both U_1 and Δ_1), then

$$K_3^J(x) = \begin{cases} -\frac{3}{2} + 4\sqrt{\frac{2}{3}}\left(1 - \sqrt{\frac{2}{3}}\,|x|\right), & |x| \leq 1 \\ 4\sqrt{\frac{2}{3}}\left(1 - \sqrt{\frac{2}{3}}\,|x|\right), & 1 < x \leq \sqrt{\frac{3}{2}} \text{ or } -\sqrt{\frac{3}{2}} \leq x < -1 \\ 0 & , \text{ elsewhere.} \end{cases}$$



For the above kernel, the estimator

$$h_3^J(n,x_0) = \frac{1}{b_3(n)} \sum_{j=1}^{n} \frac{1}{n-j+1}\, K_3^J\left(\frac{X_{(j)}-x_0}{b_3(n)}\right)$$

has the asymptotic MSE given by

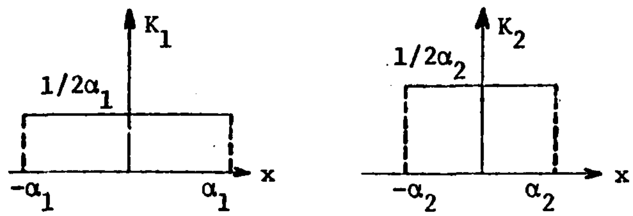
$$MSE[h_3^J(n,x_0)] \sim (1.6134)\,\frac{1}{nb_3(n)}\,\frac{h(x_0)}{1-F(x_0)} + \left[\frac{3}{56}\cdot\frac{b_3^6(n)}{6!}\,h^{(6)}(x_0)\right]^2 ,$$

and the optimal MSE is

$$MSE[h_3^J(n,x_0)]_{opt} \sim (.4725)\left[\frac{h(x_0)}{1-F(x_0)}\right]^{12/13}\left(h^{(6)}(x_0)\right)^{2/13} n^{-12/13} .$$

Example 4.2: Consider two estimators of $h(x_0)$, $h_i(n,x_0)$, $i=1,2$, based on two uniform kernels K_1 and K_2 having different bandwidths. That is, for $\alpha_1 \neq \alpha_2$

$$K_i(x) = \begin{cases} \frac{1}{2\alpha_i} , & |x| \leqq \alpha_i \\ 0 , & |x| > \alpha_i , \quad i=1,2 . \end{cases}$$



Now, $I(K_i,2) = \int u^2 K_i(u)du = \alpha_i^2/3$, and $I(K_i,4) = \int u^4 K_i(u)du = \alpha_i^4/5$, $i=1,2$. If we choose, following (4.4),

$$R = \frac{b_1^2(n)}{b_2^2(n)} \frac{I(K_1,2)}{I(K_2,2)} = \frac{b_1^2(n)\alpha_1^2}{b_2^2(n)\alpha_2^2} \neq 1 ,$$

then the new kernel, given by (4.6), will belong to A_4 .

Furthermore, if in addition to the above, we attempt to choose

$$\frac{b_1^2(n)}{b_2^2(n)} = c^2(n) = \frac{I(K_2,4)I(K_1,2)}{I(K_1,4)I(K_2,2)} = \frac{\alpha_2^2}{\alpha_1^2} ,$$

then

$$R = \frac{b_1^2(n)}{b_2^2(n)} \frac{\alpha_1^2}{\alpha_2^2} = 1 ,$$

and thus our attempt to obtain a kernel in A_6 fails.

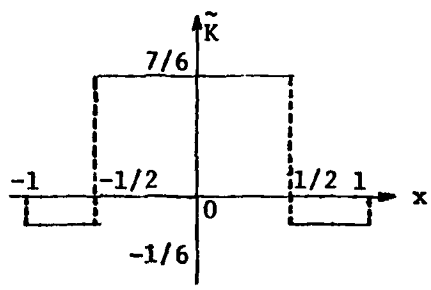
Suppose that we let $\frac{b_1(n)}{b_2(n)} = c(n) = 1$; then for $\alpha_1 > \alpha_2$, the new kernel is, by (4.6),

$$\tilde{K}(x) = \frac{K_1(x) - \frac{\alpha_1^2}{\alpha_2^2} K_2(x)}{1 - \frac{\alpha_1^2}{\alpha_2^2}}$$

$$= \begin{cases} \dfrac{(2\alpha_1)^{-1} - \alpha_1^2(\alpha_2)^{-2}(2\alpha_2)^{-1}}{1 - \alpha_1^2(\alpha_2)^{-2}} , & |x| \leqq \alpha_2 \\[2ex] \dfrac{(2\alpha_1)^{-1}}{1 - \alpha_1^2(\alpha_2)^{-2}} , & \alpha_2 < x \leqq \alpha_1 \text{ or } -\alpha_1 \leqq x < -\alpha_2 \\[2ex] 0 , & \text{otherwise.} \end{cases}$$

For example, if $\alpha_1 = 1$ and $\alpha_2 = 1/2$, then

$$\tilde{K}(x) = \begin{cases} \frac{7}{6} , & |x| \leqq \frac{1}{2} \\ -\frac{1}{6} , & \frac{1}{2} < x \leqq 1 \text{ or } -1 \leqq x < -\frac{1}{2} \\ 0 , & \text{elsewhere.} \end{cases}$$



Using this kernel, the optimal MSE of the estimator converges to zero as $n^{-8/9}$.

5. Indefinite Jackknifing and Faster Rates of Convergence

Let $h_1(n,x_0)$ be an estimator of $h(x_0)$ based on a kernel K_1 in A_2. For this estimator we know that $MSE[h_1(n,x_0)]_{opt} = O(n^{-4/5})$. Assume that $h \in C^\infty$, and for some positive real number $c \neq 1$, let

$$K_2(x) = \frac{K_1(x) - c^3 K_1(cx)}{1 - c^2} . \tag{5.1}$$

Note that the general form of K_2 is analogous to the kernel $\tilde{K}$, given by (4.6). It can be verified that $K_2 \in A_4$, since

$$\int K_2(x)dx = \frac{1}{1-c^2}\left[\int K_1(x)dx - c^3\int K_1(cx)dx\right] = \frac{1}{1-c^2}\left[1 - c^2\int K_1(u)du\right] = 1 ,$$

and

$$\int x^2 K_2(x)dx = \frac{1}{1-c^2}\left[\int x^2 K_1(x)dx - c^3\int x^2 K_1(cx)dx\right]$$

$$= \frac{1}{1-c^2}\left[\int x^2 K_1(x)dx - \int u^2 K_1(u)du\right] = 0 .$$

If $h_2(n,x_0)$ is an estimator of $h(x_0)$ based on the new kernel K_2, then from (3.14) we have $MSE[h_2(n,x_0)]_{opt} = O(n^{-8/9})$. Suppose that now we set

$$K_3(x) = \frac{K_2(x) - c^5 K_2(cx)}{1 - c^4} ; \tag{5.2}$$

then $K_3 \in A_6$. If $h_3(n,x_0)$ is an estimator of $h(x_0)$ based on K_3, then $MSE[h_3(n,x_0)]_{opt} = O(n^{-12/13})$; that is, $h_3(n,x_0)$ has a faster rate of convergence of the MSE and bias than both $h_2(n,x_0)$ and $h_1(n,x_0)$.

If we continue in this manner obtaining a kernel $K_{k-1} \varepsilon A_{2(k-1)}$, $k \geq 2$, and letting

$$K_k(x) = \frac{K_{k-1}(x) - c^{2k-1}K_{k-1}(cx)}{1 - c^{2(k-1)}} , \tag{5.3}$$

then $K_k \varepsilon A_{2k}$. If $h_k(n,x_0)$ is an estimator of $h(x_0)$ based on the kernel K_k, then we note that $MSE[h_k(n,x_0)]_{opt} = O(n^{-\frac{2(2k)}{2(2k)+1}})$. The estimator $h_k(n,x_0)$ has a faster rate of convergence of the MSE than the estimators $h_{k-1}(n,x_0),\ldots,h_1(n,x_0)$.

If this procedure is continued indefinitely, then the rate of convergence of the MSE can be brought as close to n^{-1} as is possible. That is,

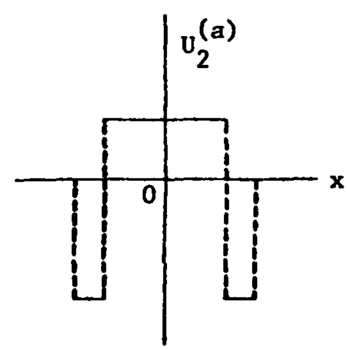
$$\lim_{k\to\infty} MSE[h_k(n,x_0)]_{opt} = O(n^{-1}) .$$

Example 5.1: Suppose that we start off with the uniform kernel $U_1 \varepsilon A_2$, $U_1(x) = \begin{cases} 1/2 , & |x| \leq 1 \\ 0 , & |x| > 1 \end{cases}$, and form a new kernel $U_2 \varepsilon A_4$ using (5.1). We shall consider the following three values of c:

(a) $c = .9$; we have

$$U_2^{(a)}(x) = \frac{U_1(x) - c^3U_1(cx)}{1 - c^2}$$

$$= \begin{cases} .7132 , & |x| \leq 1 \\ -1.918 , & 1 < x \leq \frac{1}{.9} \text{ or } -\frac{1}{.9} \leq x < -1 \\ 0 , & \text{elsewhere.} \end{cases}$$



The asymptotic MSE of $h_2^{(a)}(n,x_0) = \frac{1}{b_2(n)} \sum_{j=1}^{n} \frac{1}{n-j+1} U_2^{(a)}\left(\frac{X_{(j)}-x_0}{b_2(n)}\right)$ is

$$\mathrm{MSE}[h_2^{(a)}(n,x_0)] \sim (1.8334)\ \frac{1}{nb_2(n)}\ \frac{h(x_0)}{1-F(x_0)} + \left[(-.2460)\ \frac{b_2^4(n)}{4!}\ h^{(4)}(x_0)\right]^2 ,$$

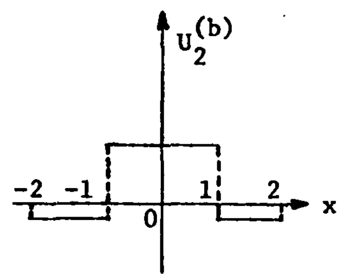
and the optimal MSE is

$$\mathrm{MSE}[h_2^{(a)}(n,x_0)]_{opt} \sim (.8779) \left[\frac{h(x_0)}{1-F(x_0)}\right]^{8/9} \left[h^{(4)}(x_0)\right]^{2/9} n^{-8/9} .$$

(b) c = .5; we have

$$U_2^{(b)}(x) = \frac{U_1(x) - c^3 U_1(cx)}{1 - c^2}$$

$$= \begin{cases} .5833 , & |x| \leq 1 \\ -.0833 , & 1 < x \leq 2 \quad \text{or} \quad -2 \leq x < 1 \\ 0 , & \text{elsewhere.} \end{cases}$$



The asymptotic MSE of $h_2^{(b)}(n,x_0) = \frac{1}{b_2(n)} \sum_{j=1}^{n} \frac{1}{n-j+1} U_2^{(b)}\left(\frac{X_{(j)}-x_0}{b_2(n)}\right)$ is

$$\mathrm{MSE}[h_2^{(b)}(n,x_0)] \sim (.3472)\ \frac{1}{nb_2(n)}\ \frac{h(x_0)}{1-F(x_0)} + \left[(-.8000)\ \frac{b_2^4(n)}{4!}\ h^{(4)}(x_0)\right]^2 ,$$

and the optimal MSE is

$$\mathrm{MSE}[h_2^{(b)}(n,x_0)]_{opt} \sim (.2600) \left[\frac{h(x_0)}{1-F(x_0)}\right]^{8/9} \left[h^{(4)}(x_0)\right]^{2/9} n^{-8/9} .$$

(c) c = .1; we have

$$U_2^{(c)}(x) = \frac{U_1(x) - c^3 U_1(cx)}{1 - c^2} = \begin{cases} .5045 , & |x| \leq 1 \\ -.0005 , & 1 < x \leq 10 \quad \text{or} \quad -10 \leq x < -1 \\ 0 , & \text{elsewhere.} \end{cases}$$

The asymptotic MSE of $h_2^{(c)}(n,x_0) = \frac{1}{b_2(n)} \sum_{j=1}^{n} \frac{1}{n-j+1} U_2^{(c)}\left(\frac{X_{(j)}-x_0}{b_2(n)}\right)$ is

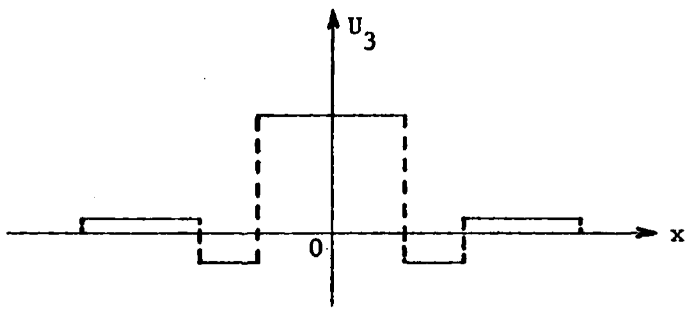
$$MSE[h_2^{(c)}(n,x_0)] \sim (.5091) \frac{1}{nb_2(n)} \frac{h(x_0)}{1-F(x_0)} + \left[(-20.000) \frac{b_2^4(n)}{4!} h^{(4)}(x_0)\right]^2 ,$$

and the optimal MSE is

$$MSE[h_2^{(c)}(n,x_0)]_{opt} \sim (.7470)\left[\frac{h(x_0)}{1-F(x_0)}\right]^{8/9}\left[h^{(4)}(x_0)\right]^{2/9} n^{-8/9} .$$

Example 5.2: If we jackknife using the kernel in Example 5.1, case (b), one more time using $c = .5$ in (5.2), we obtain

$$U_3(x) = \frac{U_2^{(b)}(x) - c^5 U_2^{(b)}(cx)}{1-c^4} = \begin{cases} .6027 , & |x| \leq 1 \\ -.1083 , & 1 < x \leq 2 \text{ or } -2 \leq x < -1 \\ .0028 , & 2 < x \leq 4 \text{ or } -4 \leq x \leq -2 \\ 0 , & \text{elsewhere.} \end{cases}$$

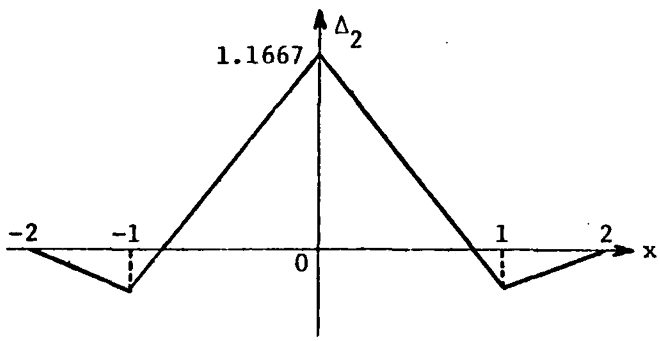


The kernel $U_3 \varepsilon A_6$, and the optimal MSE of $h_3(n,x_0)$ has a rate of convergence of

$$MSE[h_3(n,x_0)]_{opt} = O\left(n^{-12/13}\right) .$$

Example 5.3: Let us now consider a triangular kernel $\Delta_1 \varepsilon A_2$, $\Delta_1(x) = \begin{cases} 1-|x| , & |x| \leq 1 \\ 0 , & |x| > 1 \end{cases}$, and form a new kernel $\Delta_2 \varepsilon A_4$ using (5.1), with $c = .5$. We have

$$\Delta_2(x) = \begin{cases} 1.1667 - 1.25|x| , & |x| \leq 1 \\ -.1667 + .0833|x| , & 1 < x \leq 2 \text{ or } -2 \leq x < -1 \\ 0 , & \text{elsewhere.} \end{cases}$$



With this kernel, the estimator $h_2^{\Delta}(n,x_0) = \frac{1}{b_2(n)} \sum_{j=1}^{n} \frac{1}{n-j+1} \Delta_2\left(\frac{X_{(j)}-x_0}{b_2(n)}\right)$ has the asymptotic MSE given by

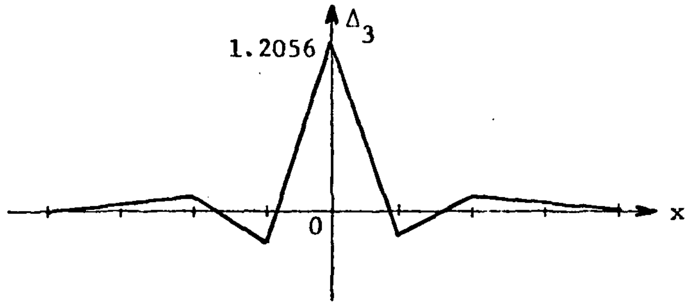
$$MSE[h_2^{\Delta}(n,x_0)] \sim (.8519) \frac{1}{nb_2(n)} \frac{h(x_0)}{1-F(x_0)} + \left[(-.2667) \frac{b_2^4(n)}{4!} h^{(4)}(x_0)\right]^2 ,$$

and the optimal MSE is

$$MSE[h_2^{\Delta}(n,x_0)]_{opt} \sim (.4522)\left[\frac{h(x_0)}{1-F(x_0)}\right]^{8/9}\left[h^{(4)}(x_0)\right]^{2/9} n^{-8/9} .$$

If we wish to obtain a kernel in A_6 using the triangular kernel Δ_1 , we let $c = .5$ in (5.2) and obtain

$$\Delta_3(x) = \frac{\Delta_2(x) - c^5\Delta_2(cx)}{1-c^4} = \begin{cases} 1.2056 - 1.3125|x| , & |x| \leq 1 \\ -.2167 + .1097|x| , & 1 < x \leq 2 \text{ or } -2 \leq x < -1 \\ .0056 - .0014|x| , & 2 < x \leq 4 \text{ or } -4 \leq x < -2 \\ 0 , & \text{elsewhere.} \end{cases}$$



The rate of the optimal MSE of the estimator $h_3^{\Delta}(n,x_0)$ based on kernel Δ_3 is of order $n^{-12/13}$.

In the following final example, we wish to estimate an exponential failure rate function at point $x_0 = 1$. To compare estimates using different kernels, the rates of convergence of the MSE's are computed. We note that the rate of convergence of the MSE is actually improved if the kernels used in estimation are not restricted to be nonnegative, although the resulting theoretical gain in efficiency may not be realized unless the sample is very large.

Example 5.4: Let the failure rate function be exponential and given by

$$h(x) = e^x , \quad 0 \leq x < \infty .$$

That is, for this form of $h(x)$, the sample $\{X_1,\ldots,X_n\}$ is from an extreme value distribution,

$$F(x) = 1 - \exp[-\exp(x)] , \quad -\infty < x < \infty .$$

Thus, for a fixed value of n and for different kernels used in the estimation, we have the following result, as given in Table 5.1.

TABLE 5.1

ESTIMATING $h(x) = e^x$ AT $x_0 = 1$ USING DIFFERENT KERNELS

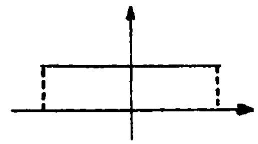
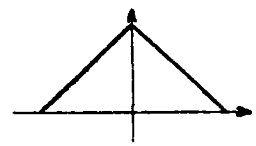
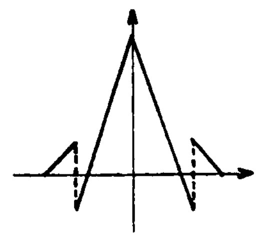
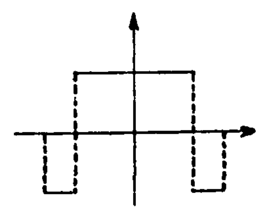
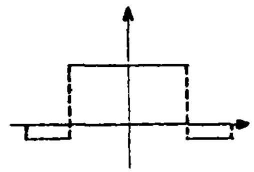
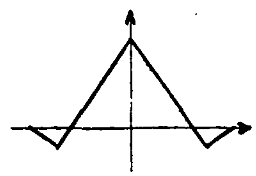
Kernel		MSE_{opt}
Uniform: $U_1(x) = \begin{cases} 1/2, & \lvert x\rvert \leqq 1 \\ 0, & \lvert x\rvert > 1 \end{cases}$		$13.5151n^{-4/5}$
Triangular: $\Delta_1(x) = \begin{cases} 1 - \lvert x\rvert, & \lvert x\rvert \leqq 1 \\ 0, & \lvert x\rvert > 1 \end{cases}$		$12.8931n^{-4/5}$
Uniform and Triangular: $K_3^J(x) = \begin{cases} -\frac{3}{2} + 4\sqrt{\frac{2}{3}}\left(1 - \sqrt{\frac{2}{3}}\lvert x\rvert\right), & \lvert x\rvert \leqq 1 \\ 4\sqrt{\frac{2}{3}}\left(1 - \sqrt{\frac{2}{3}}\lvert x\rvert\right), & 1 < x \leqq \sqrt{\frac{3}{2}} \text{ or } -\sqrt{\frac{3}{2}} \leqq x < -1 \\ 0, & \text{elsewhere} \end{cases}$		$17.0501n^{-12/13}$
Uniform and Uniform $(c = .9)$**:** $U_2^{(a)}(x) = \begin{cases} .7132, & \lvert x\rvert \leqq 1 \\ -1.918, & 1 < x \leqq \frac{1}{.9} \text{ or } -\frac{1}{.9} \leqq x < -1 \\ 0, & \text{elsewhere} \end{cases}$		$29.8784n^{-8/9}$

TABLE 5.1--*continued*

Kernel		MSE_{opt}
Uniform and Uniform $(c=.5)$: $U_2^{(b)}(x) = \begin{cases} .5833 , & \lvert x\rvert \leq 1 \\ -.0833 , & 1 < x \leq 2 \quad \text{or} \quad -2 \leq x < -1 \\ 0 , & \text{elsewhere} \end{cases}$		$8.8473n^{-8/9}$
Uniform and Uniform $(c=.1)$: $U_2^{(c)}(x) = \begin{cases} .5045 , & \lvert x\rvert \leq 1 \\ -.0005 , & 1 < x \leq 10 \quad \text{or} \quad -10 \leq x < -1 \\ 0 , & \text{elsewhere} \end{cases}$		$25.4227n^{-8/9}$
Triangular and Triangular: $\Delta_2(x) = \begin{cases} 1.1667 - 1.25 \lvert x\rvert , & \lvert x\rvert \leq 1 \\ -.1667 + .0833 \lvert x\rvert , & 1 < x \leq 2 \quad \text{or} \quad -2 \leq x < -1 \\ 0 , & \text{elsewhere} \end{cases}$		$15.3899n^{-8/9}$

Acknowledgments

The work reported here was begun when the authors were visitors at the Department of Statistics of Stanford University. We would like to thank Professor Rupert Miller for his time, and for making available to us copies of his relevant papers. Professors Donald Iglehart and Bradley Efron were kind enough to provide Man-Yuen Wong with office space and other facilities at Stanford.

REFERENCES

GRAY, H. L. and W. R. SCHUCANY (1972). *The Generalized Jackknife Statistics*. Marcel Dekker, New York.

GRENANDER, U. (1956). On the theory of mortality measurement, I and II. *Actuar. Tidskr.* **39** 90-96; 125-153.

MILLER, R. G. (1974). The jackknife--a review. *Biometrika* **61** 1-15.

MILLER, R. G. (1978). The jackknife: survey and applications. Proceedings of the 23rd Conference on the Design of Experiments in Army Research Development and Testing, ARO Report 78-2.

RICE, J. and M. ROSENBLATT (1976). Estimation of the log survivor function and hazard function. *Sankhyā* **33** 60-78.

SCHUCANY, W. R., H. L. GRAY, and D. B. OWEN (1971). On bias reduction in estimation. *J. Amer. Statist. Assoc.* **66** 524-533.

SCHUCANY, W. R. and J. P. SOMMERS (1977). Improvement of kernel type density estimation. *J. Amer. Statist. Assoc.* **72** 420-423.

SETHURAMAN, J. and N. D. SINGPURWALLA (1978). Large sample estimates and uniform confidence bounds for the failure rate function based on a naive estimator. FSU Statistics Report M472, The Florida State University.

SINGPURWALLA, N. D. and M.-Y. WONG (1980). Fourier integral estimate of the failure rate function and its mean square error properties. Technical Paper Serial T-416, Institute for Management Science and Engineering, The George Washington University.

SHAPIRO, H. S. (1969). *Smoothing and Approximation of Functions*. Van Nostrand Reinhold, New York.

WATSON, G. S. and M. R. LEADBETTER (1964a). Hazard analysis I. *Biometrika* **51** 175-184.

WATSON, G. E. and M. R. LEADBETTER (1964b). Hazard analysis II. *Sankhyā* **26** 101-116.